UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

INVENTORY ROUTING : REDUCTION FROM AN ANNUAL TO SHORT PERIOD PROBLEM

by

M. Dror* and M. Ball**

* Department of Industrial Engineering and Management,
Ben-Gurion University of the Negev, Beersheva, Israel,
Visiting Professor, Centre de recherche sur les transports,
Université de Montréal, Summer 1984.

** College of Business and Management,
University of Maryland at College Park.

Université de Montréal
Centre de recherche sur les transports - Publication #353
Août 1984

ABSTRACT

The inventory routing problem involves a set of N customers, where each customer has a different demand on each day. Given a central supplier, the objective is to minimize the annual delivery costs while attempting to insure that no customer runs out of the commodity at any time. Assuming that for the N customers their daily consumption rates are each approximately normally distributed, we develop and test a customer selection procedure that reduces an inventory routing problem over an annual time base to a sequence of short period deterministic optimization problems.

- - - - - - - - - - -

PROBLEM DESCRIPTION

The inventory routing problem involves a set of N customers, where each customer has a different demand on each day. For example, each customer utilizes a commodity such as heating oil or methane at an estimated consumption rate. Each customer possesses a known capacity (for example, the size of his tank to hold home heating oil). The objective is to minimize the annual delivery costs while attempting to insure that no customer runs out of the commodity at any time. The impetus behind this distribution system is the importance of maintaining a sufficient supply of inventory at a customer's location. Each customer consumes more than his fixed tank capacity can hold during the course of a year ; thus, in order not to run out of the commodity, a customer requires a number of replenishments annually. Though the objective is to minimize the annual delivery cost, the operational time horizon for which the problem can be formulated and solved is of much shorter duration, say, of a week or two weeks for the following reasons :

a) The validity of a formulation with an annual time base is questionable because some parameter values are uncertain over that long a period, and others change.

b) The number of constraints and variables for such a formulation is prohibitively high.

The procedure of reducing an annual optimization problem and selecting the set of customers actually replenished in a short-operational time period is of major importance in developing a detailed solution to the inventory routing problem. In this paper we present a reduction and customer selection procedure. Given a data base, conditons are developed which enable a transition from a stochastic problem (i.e., customer's daily consumption rate follows a certain distribution) to deterministic problem in which customer's daily demand is treated as known. Throughout the analysis of

this problem we assume that if a customer runs out of inventory, he is replenished immediately by an emergency service. On a given day the customer's demand is defined as the capacity of the tank minus the stock on hand. It implies a replenishment policy under which a customer's tank is always filled up when he is serviced. Consequently, if the customers consumption rate is known, the amount delivered is determined by the day of delivery.

In the process of developing problem reduction and customer selection procedure, we analyze different terms associated with a cost of replenishing a customer on a given day. We introduce a concept of the expected annual incremental cost of replenishing a customer on a given day and develop a mathematical expression for this cost. Thus, for each customer we obtain a replenishment "cost" sequence.

An important characteristic of this sequence is its unimodality. We prove the unimodality property and based on it determine the last day by which the customer can be scheduled for replenishment.

When reducing an inventory distribution problem over an annual time horizon to a sequence of optimization problems over a weekly time horizon a proper reduction procedure is of major importance. Following Fisher et al, [7] we claim that unless incentives are provided for the amount delivered, an optimization model formulated for a fixed weekly period would make minimal deliveries within a single period, deferring as many deliveries as possible to later planning periods. If there is no incentive to make larger deliveries to a given customer, this reduces the cost for the current planning period, but at the expense of future planning periods. The result could be a very high annual direct cost because of many small deliveries.

In pure inventory problems where the objective is to minimize the average cost over a long time horizon, the model usaully contains some fixed costs associated with each delivery (or associated with the beginning

of an inventory build up). The relative size of these fixed costs will have a significant impact on the size of the optimal inventory deliveries. In general, the higher the fixed costs the larger the volumes delivered each time in order to reduce the number of deliveries during the time horizon (i.e., reduce the total fixed cost associated with the deliveries). In the IRP, the fixed cost associated with a delivery to a customer is expressed as the direct added cost of this customer's inclusion in the distribution. For the same reasons as in the pure inventory model, the customer's share in the distribution cost and the value of this share impact directly on the optimal volume delivered. (A delivery always fills up the tank; thus the optimal volume delivered is determined by the day of delivery.)

An extreme example would be the case in which the fixed delivery cost is zero : In this case the distributor would make frequent deliveries exactly matched to the customer's needs. Water, gas and electricity are delivered in this way in our society because in effect the fixed cost per delivery is zero.

In this paper we attempt to explore the relationship between the annual distribution cost, the fixed delivery cost, and the amount delivered, and to formulate a mathematical model which expresses this effectively. More specifically, based on these costs, we develop customer selection strategies which will determine the customers replenished on each given day.

Two different customer selection procedures are presented. The procedure which we describe first was actually tested on our data base from York, Penna. district for twelve consecutive weekly planning periods. We refer to this procedure as the actual customer selection procedure. The second customer selection procedure described here represents an alternative selection process with a different approach to the selection problem.

This alternative customer selection process was not tested on our actual data base. As such, we cannot assess its comparative performance, but only qualify its advantages and weaknesses. We refer to the second customer selection process as the alternative procedure.

1. Computation of Cost Coefficients c_{ik}

Assume that we know each customer's expected daily consumption rates In practice, the expected daily consumption rate is estimated by analyzing the consumption history. (See for example [1] and [2]. Some notation follows :

μ_{ik} - denotes the expected consumption rate for customer i on day k.

d_{ik} - denotes the expected delivery size (volume) required to fill the tank of customer i on day k. It is referred to as the expected demand.

A_i - denotes the expected annual consumption for customer i.

ND - denotes the number of days in the planning period and is determined in advance.

N - denotes the number of customers ($\{N\}$ denotes the set of the customers.) We assume that N is fixed.

With each customer we associate a sequence of daily demands. In between replenishments the expected daily demands for customer i form a monotonically nondecreasing sequence $\{d_{ik}\}_{k=1}^{K_i}$ (i.e., $0 < d_{i1} < d_{i2} < \ldots < d_{iK_i}$), where K_i denotes the last day before customer i is expected to stock-out if no delivery has reached him on day $k < K_i$. Let p_{ik} be the number of times in the course of one year (i.e., 365 days as of today) that the i^{th} customer has to be replenished if resupplied when his expected demand exactly matches the expected demand on day k. Assume that A_i, the expected annual consumption for customer i, is known. Then p_{ik} is calculated as follows : $p_{ik} = \frac{A_i}{d_{ik}}$. Note that d_{ik} in the formula is not considered a random variable but is given an appropriate value for each day k.

The p_{ik}'s form a monotonically nonincreasing sequence for each i since, all other factors being equal, if a customer is replenished when his expected demand is greatest, he will require fewer replenishments annually (i.e., $p_{i1} > p_{i2} > \ldots > p_{iK_i}$). As an example, if a customer whose tank has a 25 unit capacity, and who consumes 100 units annually has a $d_{ik} = 25$, then $p_{ik} = 4$ (i.e., 4 visits a year). On the other hand if $d_{ik} = 5$, then $p_{ik} = 20$ (20 visits a year).

In our analysis of the IRP, it is essential to identify the behavior of finite and infinite sequences. As we can observe so far, each customer's expected daily demands and the p_{ik}'s are finite sequences. Later, we develop an expression to calculate our estimate of the cost of scheduling a customer for replenishment on a specific day. This calculation constructs another finite sequence which we examine for its minimal value. A minimal (maximal) value of a sequence is important, but so is the relationship among a sequence values in the "neighborhood" of the extreme point. It may indicate in our case the day on which a customer will be replenished, together with the cost of deviating from the scheduled day. In order to facilitate the analysis of finite or infinite sequences we define the notion of a "convex sequence".

Definition: Given a finite or infinite sequence $\{x_i\}$, $i \in I$, we say that the sequence $\{x_i\}$, $i \in I$ is convex if for all $i \in I$, $x_{i-1} - x_i > x_i - x_{i+1}$. The sequence $\{x_i\}$, $i \in I$ is strictly convex if $x_{i-1} - x_i > x_i - x_{i+1}$ for all $i \in I$.

(We can define a concave sequence similarly.)

Claim 1.1: If $d_{i,k+1} - d_{ik} = d_{ik} - d_{i,k-1} > 0$ for all $k = 1,\ldots,K_i - 1$ (i.e., if customer i's expected consumption rate μ_{ik} is positive and equal for all $k = 1,\ldots,K_i$) then the sequence $\{p_{ik}\}_{k=1}^{K_i}$ is decreasing and strictly convex.

Proof: Denote by $\mu_i = d_{i,k+1} - d_{ik}$. Using the formula to calculate p_{ik} (i.e., $p_{ik} = \frac{A_i}{d_{ik}}$) we get that

$$p_{i,k-1} - p_{ik} = \frac{A_i}{d_{i,k-1}} - \frac{A_i}{d_{ik}} = \frac{A_i}{(d_{ik}-\mu_i)} - \frac{A_i}{d_{ik}} = \frac{A_i\mu_i}{(d_{ik}-\mu_i)d_{ik}} = \frac{A_i\mu_i}{d_{i,k-1}d_{ik}}$$

similarly

$$p_{ik} - p_{i,k+1} = \frac{A_i\mu_i}{(d_{i,k+1}-\mu_i)d_{i,k+1}} = \frac{A_i\mu_i}{d_{ik}d_{i,k+1}}.$$

According to the definition of a strictly convex sequence we have to prove that

$$\frac{A_i\mu_i}{d_{i,k-1}d_{ik}} > \frac{A_i\mu_i}{d_{ik}d_{i,k+1}} \quad \text{or} \quad \frac{d_{i,k+1}}{d_{i,k-1}} > 1.$$

But from the definition of μ_i

$$\frac{d_{i,k+1}}{d_{i,k-1}} = \frac{d_{i,k-1}+2\mu_i}{d_{i,k-1}} = 1 + \frac{2\mu_i}{d_{i,k-1}}$$

Given that $\mu_i > 0$, and assuming finite demands, $\frac{2\mu_i}{d_{i,k-1}}$ is a positive number. Consequently $1 + \frac{2\mu_i}{d_{i,k-1}} > 1$.

Q.E.D.

The distribution cost of replenishing customer i consists of the cost of arriving at the customer's location (cost A) plus the cost of unloading the commodity (cost B). Cost A must be calculated differently for two different scenarios: One cost, denoted by L_i, corresponds to the cost of arriving at customer i's location on a regular route. (This indicates that the customer did not stock-out yet.) L_i depends of course on the structure of the particular route in which the customer i takes part. The other cost, denoted by S_i, corresponds to the cost of arriving at customer i's location on a special delivery route. (This indicates that the customer did stock-out, and a special delivery was necessary.) In an extreme

case, S_i equals the cost of a round trip from the depot to customer i. We present the description of several methods we have developed to estimate the cost L_i in Appendix I. For our present focus, we can assume that L_i's are given for all customers, and we concentrate on the coefficients c_{ik}.

The coefficients c_{ik} represent an estimate of the annual incremental cost of scheduling customer i for replenishment on a regular delivery route when his expected demand consistently matches the demand d_{ik}.

Denote by f the fixed cost associated with each delivery. Note that f is a constant, independent of the demand volume, day, and customer location. For each demand level d_{ik}, $i = 1,\ldots,N$ and $k = 1,\ldots,K_i$ we compute the annual cost of arriving at customer i's location when his demand matches d_{ik}. This cost is represented by the sequence $\{p_{ik}(L_i + f)\}_{k=1}^{K_i}$.

For any given day k, if throughout the year we consistently replenish customer i when his expected demand matches d_{ik}, then we will be replenishing this customer p_{ik} times; each time incurring a cost that equals the cost of arriving at customer i's location, plus the fixed cost associated with each delivery. To shorten the notation set, $p_{ik}(L_i + f) \equiv \bar{c}_{ik}$. For each customer, we have a finite nonincreasing sequence that reaches its minimal value at its last member $\bar{c}_{iK_i}$. Note that there is no need to add the cost of unloading each delivery, because we assume that the customer's annual demand A_i will always be satisfied. Thus, the annual unloading cost is a constant independent of the distribution system.

Define for each i and k ($i = 1,\ldots,N$; $k = 1,\ldots K_i$) the cost coefficients c_{ik} as $c_{ik} \equiv \bar{c}_{ik} - \bar{c}_{iK_i}$. (Note that $c_{iK_i} \equiv 0$.)

The c_{ik}'s correspond to the intuitive notion that if the company has to provide the same type of service to a customer more often than absolutely necessary, the cost of this service will rise. Note that if we

consistently replenish customer i just as the customer is stocking out, the incremental cost of this replenishment policy is zero. Next, we prove a corollary which enables us to derive an important property for the expected incremental cost sequence of scheduling a customer for replenishment.

Corollary 1.1: Id $d_{i,k+1} - d_{ik} = d_{ik} - d_{i,k-1} > 0$ for all $k = 1,\ldots,K_i-1$ then the sequence $\{c_{ik}\}_{k=1}^{K_i}$ is decreasing and strictly convex.

Proof: It is clear that it is a decreasing sequence ; we prove that it is also strictly convex.

$$c_{ik} = \bar{c}_{ik} - \bar{c}_{iK_i} = p_{ik}(L_i + f) - p_{iK_i}(L_i + f) = (L_i + f)(p_{ik} - p_{iK_i})$$

$$c_{i,k-1} - c_{ik} = (L_i + f)(p_{i,k-1} - p_{iK_i} - p_{ik} + p_{iK_i}) = (L_i + f)(p_{i,k-1} - p_{ik})$$

In Claim 1.1, we proved that $\{p_{ik}\}_{k=1}^{K_i}$ is a decreasing and strictly convex sequence (i.e., $p_{i,k-1} - p_{ik} > p_{ik} - p_{i,k+1}$) thus:

$$c_{i,k-1} - c_{ik} = (L_i+f)(p_{i,k-1}-p_{ik}) > (L_i+f)(p_{ik}-p_{i,k+1}) = c_{ik} - c_{i,k+1}.$$

Q.E.D.

2. Analysis of the Expected Cost of Assigning Customer i for Replenishment on Day k

We denote by r_{ik} the daily consumption rate for customer i on day k. Assume that r_{ik}'s for each i are independent, identically distributed normal random variables with known parameters (i.e., $\mu_{ik} = \mu_i$ and $\sigma_{ik} = \sigma_i$ for all $k = 1,\ldots,K_i$). We denote for customer i the cost of stocking out by S_i. Assume that S_i depends only on i and is not a function of the day. This situation corresponds to one in which a special delivery will take place to replenish the customer in the event of stock-out. Denote by I_i customer i's tank capacity. Denote by $P_i(k)$ the probability that customer i will stock-

out exactly on day k. We let $f(x|k\mu_i, \sqrt{k}\,\sigma_i)$ denote the normal density function with mean $k\mu_i$ and variance $k\sigma_i^2$.

Let $c^i = \frac{\sigma_i}{\mu_i}$ be the coefficient of variation.

Assume that the cost coefficients c_{ik} are given for all $i = 1,\ldots,N$ and $k = 1,\ldots,K_i$. This means that we know the incremental cost of delivery schedules for all the customers on each day of the planning period. In the IRP, on every delivery a customer's tank is filled. Stock-out will occur when the customer's tank becomes empty before a scheduled delivery.

The probability of a stock-out exactly on day k for customer i is given by the following expression :

$$P_i(k) = \text{Prob.}\left\{\sum_{t=1}^{k-1} r_{it} < I_i < \sum_{t=1}^{k} r_{it}\right\} = \text{Prob.}\left\{\sum_{t=1}^{k} r_{it} > I_i \middle| \sum_{t=1}^{k-1} r_{it} < I_i\right\} \bullet$$

$$\text{Prob.}\left\{\sum_{t=1}^{k-1} r_{it} < I_i\right\} = \text{Prob.}\left\{\sum_{t=1}^{k-1} r_{it} < I_i\right\} - \text{Prob.}\left\{\sum_{t=1}^{k} r_{it} < I_i\right\}$$

$$= \int_{-\infty}^{I_i} f(x|(k-1)\mu_i, \sqrt{k-1}\,\sigma_i)\,dx - \int_{-\infty}^{I_i} f(x|k\mu_i, \sqrt{k}\,\sigma_i)dx$$

At this point we present and prove a Lemma that plays a key role in our analysis of the expected cost of assigning a customer for replenishment on a specific day. For each customer, denote by w_i the real number that satisfies the following equality : $1 + w_i = \frac{I_i}{\mu_i}$.

Lemma 2.1: Given the case where the coefficient of variation c^i is less than or equal to one, the probability $P_i(m)$ that a customer i stocks-out exactly on day m is increasing as a function of m for all $m < w_i$.

Proof[1]: In the proof we drop the index i to simplify notation.

Denote by $F(m) + \int_{-\infty}^{I} f(x|m\mu, \sqrt{m}\,\sigma)dx$.

To prove that P(m) is an increasing function of m as m increases, is the same as proving that -P(m) decreases as m increases. For -P(m) to decrease as m increases, the difference F(m) - F(m-1) has to decrease. If $\frac{d}{dm}$ F(m) decreases, this is a sufficient condition for the difference F(m) - F(m-1) to decrease. In other words, if $\frac{d}{dm}$ F(m) decreases it means that there exists a sufficiently small h such that $\frac{F(m+1+h - F(m+1)}{h} < \frac{F(m+h) - F(m)}{h}$ or F(m) - F(m-1) decreases as m increases ($m\varepsilon(0,w]$).

An equivalent sufficient condition is obtained if we prove for $m\varepsilon(0,w]$, that $\frac{d^2}{dm^2}$ F(m) is negative. (A negative second derivative implies that the first derivative decreases.) We write this condition explicitly and prove that it holds in our case.

$$(2.1)\quad \frac{d^2}{dm^2} \int_{-\infty}^{I} f(x|m\mu,\ \sqrt{m}\,\sigma)dx < 0$$

First, a basic transformation : $\int_{-\infty}^{I} f(x|m\mu,\ \sqrt{m}\,\sigma)dx = \frac{1}{2} \int_{-\infty}^{z} e^{-x^2/2}dx$

where $z = \frac{I - m\mu}{\sqrt{m}\,\sigma}$.

Let $g(z) = \int_{-\infty}^{z} e^{-x^2/2}dx$, and differentiate g(z) twice with respect to m.

$$\frac{d}{dm} g(z) = e^{-z^2/2} \frac{dz}{dm} \quad \text{and} \quad \frac{d^2}{dm^2} g(z) = -ze^{-z^2/2} \left(\frac{dz}{dm}\right)^2 + e^{-z^2/2} \frac{d^2z}{dm^2}$$

$$\frac{dz}{dm} = \frac{-(I+m\mu)}{2m^{3/2}\sigma} \quad \text{and} \quad \frac{d^2z}{dm^2} = \frac{3I+m\mu}{4m^{5/2}\sigma} \quad .$$

---

[1] I would like to thank Filip Gheysens for his help in proving Lemma 2.1.

The sufficiency condition (2.1) is now expr4essed as follows :

$$\frac{d^2}{dm^2} g(z) = - \frac{(I-m_\mu)(I+m_\mu)^2}{m^{1/2}\sigma\, 4m^3\, \sigma^2} e^{-z^2/2} + \frac{(3I+m_\mu)}{4m^{5/2}\sigma} e^{-z^2/2} < 0.$$

We multiply this inequality by $\frac{4\sigma m^{5/2}}{e^{-z^2/2}}$, which is always positive. The condition (2.1) has now the following form :

$$\frac{-(I - m_\mu)(I + m_\mu)^2}{m\sigma^2} + (3I + m_\mu) < 0$$

or

$$3mI\sigma^2 + m^2\mu\sigma^2 - (I^2 - m^2\mu^2)(I + m_\mu) < 0. \qquad (2.2)$$

Substitute I by $(w+1)\mu$ in the inequality (2.2). Remember, we picked w such that $(w + 1)\mu = I$). The inequality (2.2) has now the following form :

$$3m(w+1)\sigma^2\mu + m^2\sigma^2\mu - (w+1)m^3\mu^3 - (w+1)^2m^3 + (w+1)m^2\mu^3 + m^3\mu^3 < 0.$$

Divide this inequality by μ^3 and substitute the coefficient of variation c for $\frac{\sigma}{\mu}$. The sufficiency condition is now of the following form :

$$3(mwc^2-w^2) + (3mc^2-w) + (m^2w-mw^2) + (m^2(c^2+1)-2mw) + (m^3-w^3) - (m+1) < 0.$$

This negative inequality holds for all $m < w$ given that $c < 1$ with one negative term $-(m+1)$ to spare. It proves that our sufficiency condition holds, i.e. our function P(m) increases as m increases for all $m \in (0,w]$.

Q.E.D.

$P_i(m)$ is the probability that customer i stocks-out on day m. The probability that customer i will stock-out before day k is given by $\sum_{t=1}^{k-1} P_1(t)$.

Now we are able to present the expression for the expected annual incremental cost of scheduling a delivery to customer i on day k, understanding the behavior of all the terms involved. This expression has the following form :

$$\sum_{t=1}^{k-1} (S_i+c_{it})P_i(t) + (1-\sum_{t=1}^{k-1} P_i(t))c_{ik} = \sum_{t=1}^{k-1} (S_i+c_{it}-c_{ik})P_i(t) + c_{ik}.$$

First we interpret the expression $\sum_{t=1}^{k-1} (S_i+c_{it})P_i(t)$.

$(S_i+c_{it})P_i(t)$ is the expected annual incremental cost for customer i, given that a stock-out occurred exactly on day t. S_i is the cost of special delivery and c_{it} is the annual incremental cost of replenishing customer i consistently when his demand matches the demand that was expected on day t. Keep in mind that we always replenish the customer to his full tank capacity. $\sum_{t=1}^{k-1} (S_i+c_{it})P_i(t)$ represents the expected annual incremental cost if the customer stocks-out before day k.

$(1-\sum_{t=1}^{k-1} P_i(t))$ is the probability that customer i is replenished on day k as scheduled. Multiply this by the coefficient c_{ik}, and then it expresses the expected annual incremental cost of replenishing customer i on day k as scheduled.

Denote the term $\sum_{t=1}^{k-1}(S_i+c_{it}-c_{ik})P_i(t)$ by $W_i(k)$. Note that $c_{it}-c_{ik} > 0$ because of the strict convexity of the sequence $\{c_{ik}\}_{k=1}^{K_i}$ (Corollary 2.1). Then, $\{W_i(k)\}_{k=1}^{K_i}$ is a strictly convex and increasing sequence as a consequence of Lemma 2.1. Denote by $E_i(k) = W_i(k) + c_{ik}$ and set $P_i(0)$ as zero for all i. $\{E_i(k)\}_{k=1}^{K_i}$ is a finite sequence for each i and is formed as an additive composition of two finite sequences $\{W_i(k)\}_{k=1}^{K_i}$ and $\{c_{ik}\}_{k=1}^{K_i}$. Both

of these sequences are strictly convex. In addition, the first sequence is increasing and the second is decreasing.

Theorem 2.1: The sequence $\{E_i(k)\}_{k=1}^{K_i}$ is unimodal and strictly convex.

Proof: We prove that the sequence is strictly convex. Obviously the unimodality follows.

For the sequence $\{E_i(k)\}_{k=1}^{K_i}$ to be strictly convex, the following inequality has to hold :

$$E_i(k-1) - E_i(k) > E_i(k) - E_i(k+1) \quad \text{for all } k < K_i - 1$$

It is the same as the following inequality:

$$W_i(k-1) + c_{i,k-1} - W_i(k) - c_{ik} > W_i(k) + c_{ik} - W_i(k+1) - c_{i,k+1}$$

or

$$W_i(k-1) - W_i(k) + c_{i,k-1} - c_{ik} > W_i(k) - W_i(k+1) + c_{ik} - c_{i,k+1}.$$

From Corollary 2.1, we know that $c_{i,k-1} - c_{ik} > c_{ik} - c_{i,k+1}$ and as a consequence of Lemma 2.1 we have that $W_i(k-1) - W_i(k) > W_i(k) - W_i(k+1)$. Thus the inequality holds.

Q.E.D.

Before exploiting the implications of Theorem 2.1, we discuss the normality assumptions for the daily consumption rate for customer i on day k (i.e., the r_{ik}). Essentially, we cannot claim that a customer's daily consumption rate is normally distributed. The Gaussian curve has a tail that corresponds to negative values for the variable which vary from -∞ to 0 and which imply a negative consumption rate ; obviously this cannot be. Thus what we may assume is that the daily consumption rate approximates a truncated normal distribution. In Lemma 2.1 we are examining the sums of

independent, identically distributed random variables which correspond to the daily demand and not the daily consumption rate. In the case where the stardard deviation of r_{ik}'s is substantially smaller than their mean, our truncated normal distribution is approximately normal and by invoking the central limit theorem, the sums of the r_{ik}'s come even closer to a normal distribution.

When we examine the data on which this study was tested, most of the mean daily consumption rates far exceed their corresponding stardard deviations. In Lemma 2.1, the extreme case is where the mean for the daily consumption rate and the standard deviation are the same. In this case too, the central limit theorem ensures approximate normality of the sums of the r_{ik}'s but it may require a large number of terms in the summation. In [8] Francis describes the behavior of the sums of two and three identically distributed truncated normal random variables. Though he does not draw conclusions as to how fast sums approximate a normal distribution, by examining the relationship between the sums of truncated normal vs. normal distribution, we conjecture that the result of Lemma 2.1 holds for truncated normal distribution (as long as the standard deviation does not exceed the mean for the single variable).

The most important result of Theorem 2.1 is the unimodality of the sequence $\{E_i(k)\}_{k=1}^{K_i}$. In essence, this implies that at one point the expected incremental cost of scheduling customer i for replenishment is minimal. We denote this point by k*. That is, $E_i(k^*) = \min_k E_i(k)$.

To find the point k* we first identify the smallest k such that $(S_i+c_{i,k-1}-c_{ik})P_i(k) > c_{ik}-c_{i,k+1}$. Denote this k by $\bar{k}$. In order to identify k*, we examine the value of $E_i(k)$ for three points ; $\bar{k}-1, \bar{k}, \bar{k}+1$. The $k\in\{\bar{k}-1,\bar{k},\bar{k}+1\}$ for which $E_i(k)$ is minimal is the k*. This concludes our analysis of the sequence $\{E_i(k)\}_{k=1}^{K_i}$.

3. Safety Stock Level

We are now ready to define and identify the safety stock level for each customer. The concept of a safety stock level as defined in the literature, is slightly different from the one we will employ. For example, Thomopoulos in [11] describes safety stock as follows : "The safety stock is planned only as a precautionary measure and is an excess amount of stock that is available to meet higher than average demands." Laaks in his M.Sc. thesis [10] defines the safety stock as : "A buffer allowance designed to insulate inventory against the variability and uncertainty inherent in item demand...". These two authors, and others, use the term safety stock in connection with the inventory service level which equals the filled demand divided by the total demand. In the IRP we do not have a service level concept because the total demand is always filled.

For our purpose, then, the safety stock level for customer i is defined as the expected stock level for this customer's day k*. This means that the safety stock level corresponds to the stock level expected on the day the replenishment cost is minimal.

The ideal replenishment scheduling policy would be to service the customer always on his day k*. If the fleet's capacity always matched the demand of the customers scheduled for replenishment by this ideal policy, then this would be the best policy. In practice, however, the fleet is fixed and the demand fluctuates.

It is interesting to note that the safety stock level as defined here will change according to a customer's expected consumption during the relevant planning periods. At different times of the year, the same customer may have different safety stock levels.

Figures 3.1 to 3.4 illustrate the probability of a customer stocking out, the convexity of the sequences of the c_{ik} coefficients and the $E_i(k)$'s and the corresponding k*'s for two customers. In addition, we obtain the safety stock levels.

In Figure 3.1 we plot the expected annual incremental cost of replenishing a specific customer, customer number 3, as a function of the day of replenishment. This expected annual incremental cost is denoted by $E_3(k)$. We have also plotted in Figure 3.1 the values of the cost coefficients c_{3k} as a function of the day. In Figure 3.3 we repeated the same plots for a different customer, customer number 6, whose average daily consumption rate is about twice that of the customer plotted in Figure 3.1.

In Figure 3.2 we present the plot of the probability that customer number 3 will stock out before day k, as a function of the day. The same plot is repeated for customer number 6 in Figure 3.4 These figures represent the actual analysis of the data for two customers from York, PA, for the month of January.

4. Customer Selection Process

The purpose of analyzing the behavior of the $\{E_i(k)\}_{k=1}^{K_i}$ sequence, and calculating the safety stock level for each customer, is to facilitate selection of customers to be replenished during a given planning period.

First, we justify and state a selection principle that is adopted in the customer selection process.

The sequence $\{E_i(k)\}_{k=1}^{K_i}$ represents the expected incremental annual cost for scheduling customer i for replenishment on day k. This sequence is unimodal and its minimum is attained on day k*. For any $k > k^*$, $\sum_{t=1}^{k-1} P_i(t)$, which is the probability of customer i stocking out before day k, is

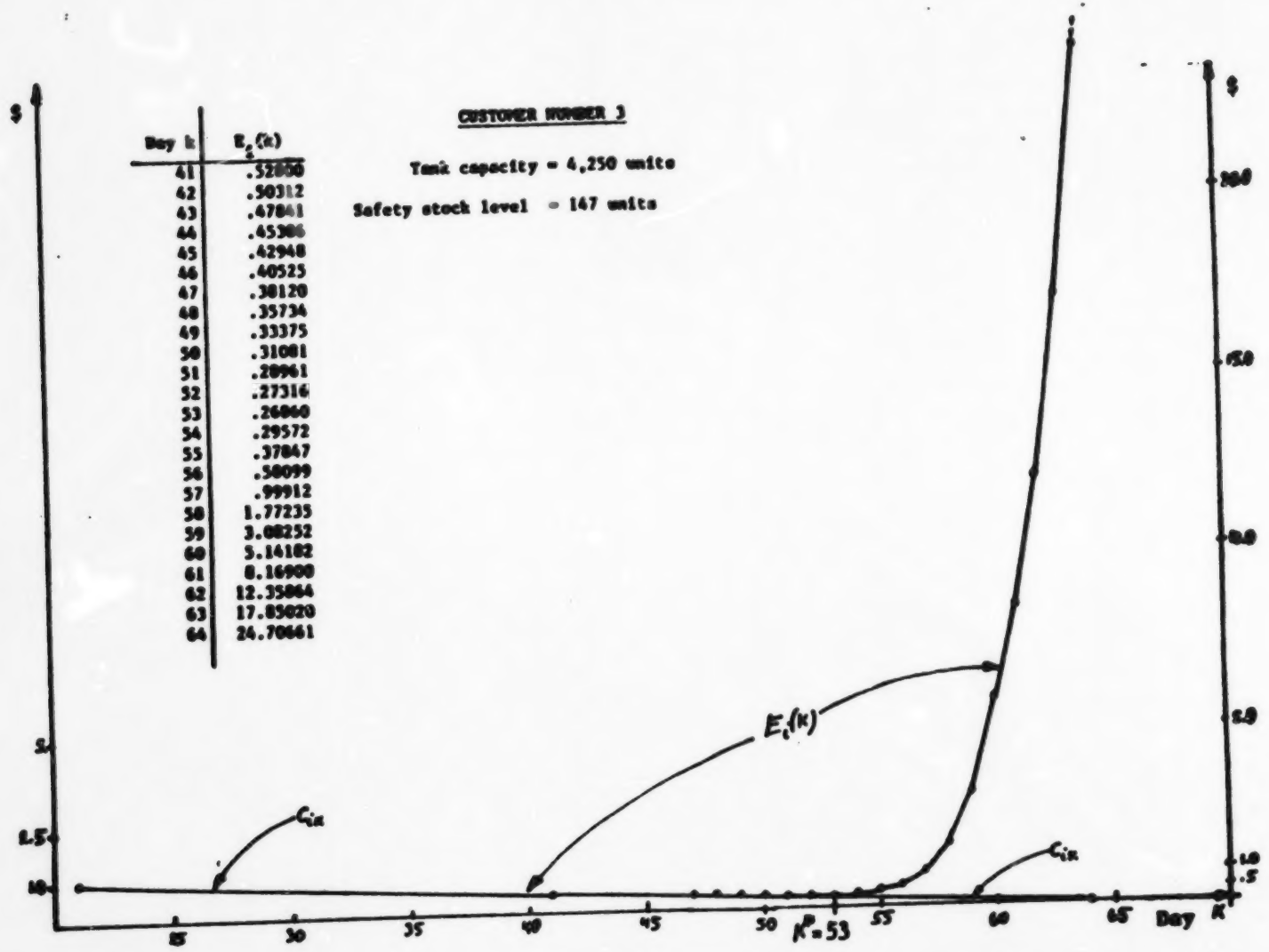


Figure 3.1 : The graph and the table present the expected annual incremental cost $E_3(k)$ of replenishing customer number 3 as a function of the scheduled day of replenishment. We also plotted on this graph the cost coefficients c_{3k}, for customer number 3, as a function of the day.

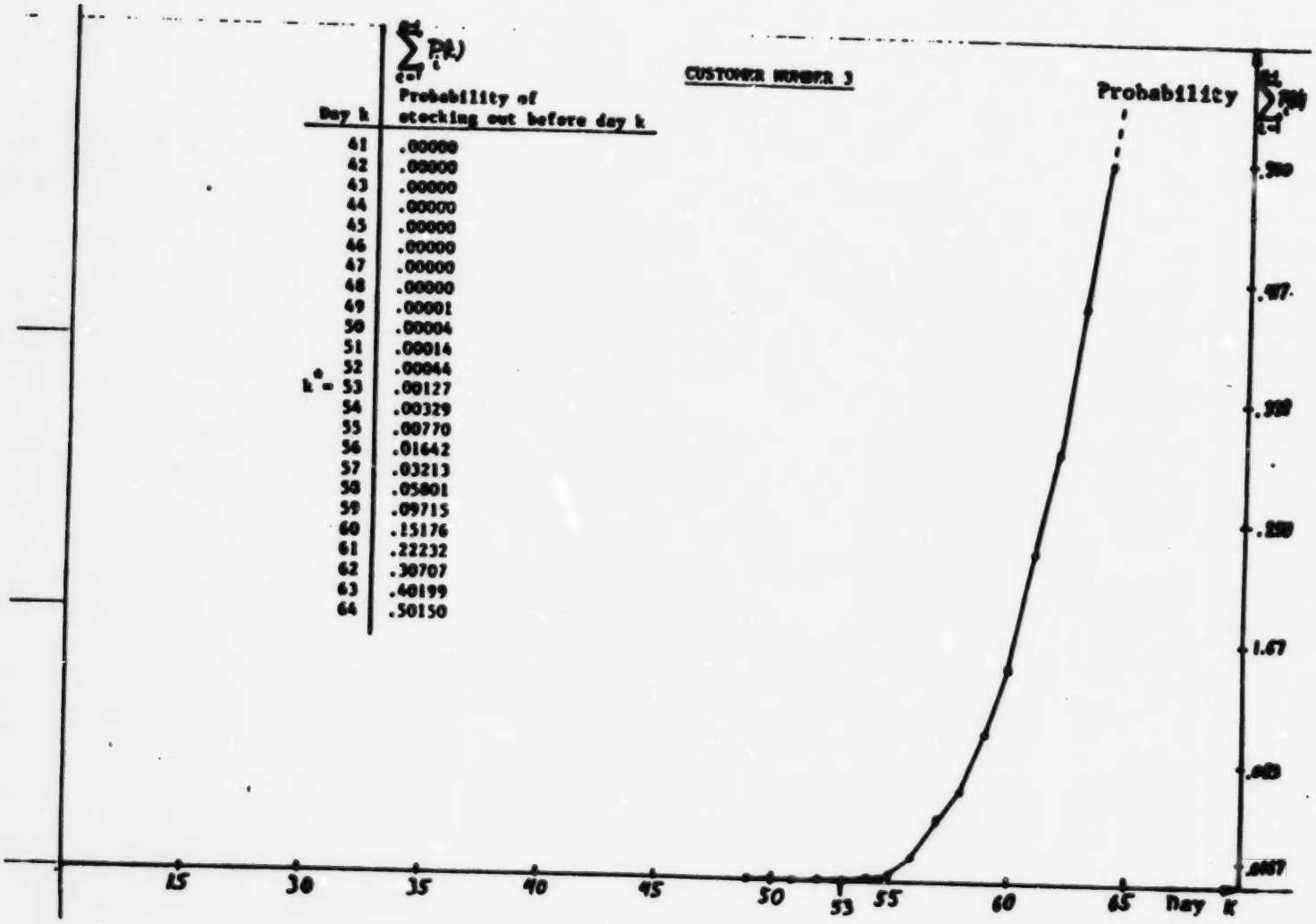
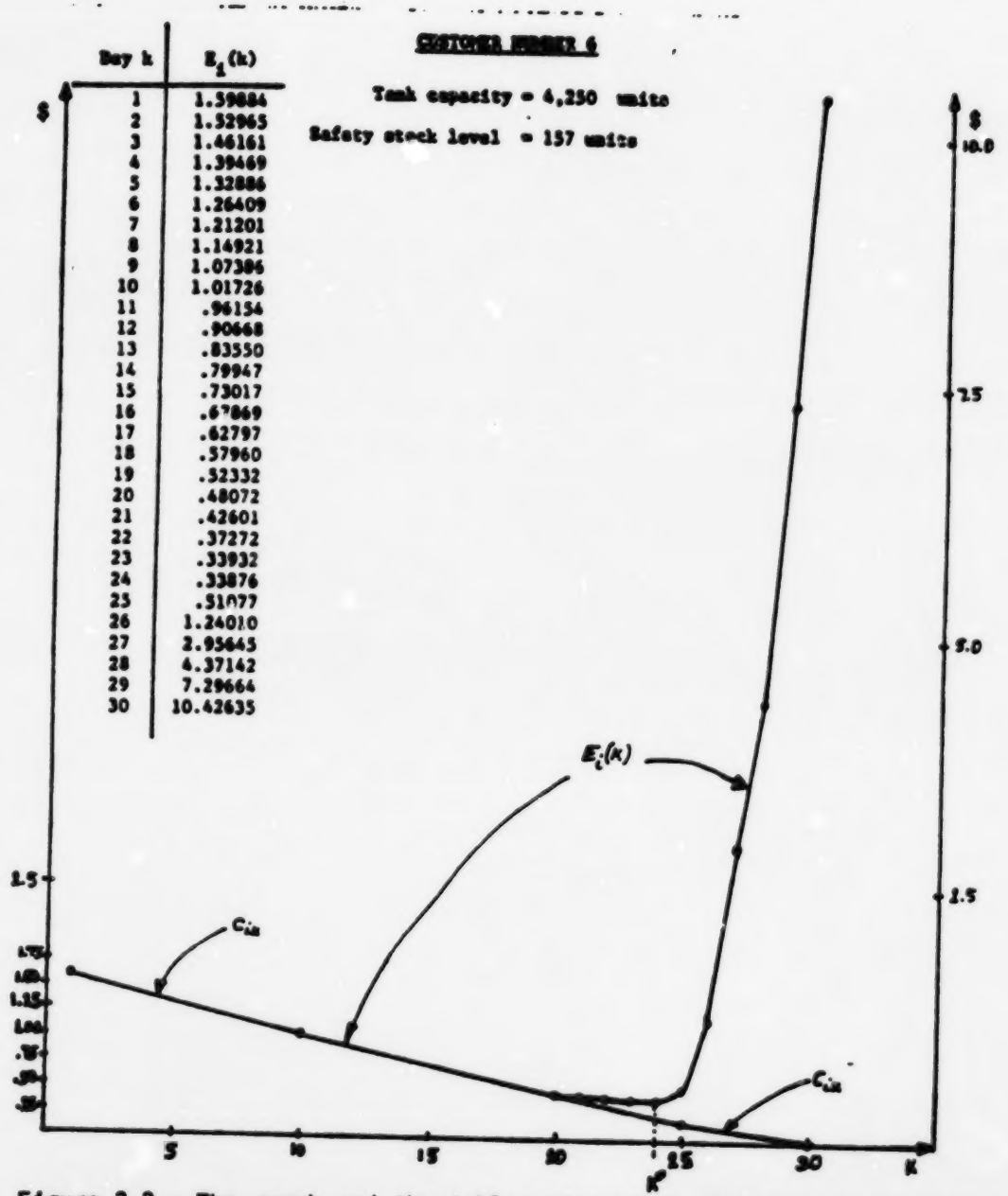


Figure 3.2 : The graph and the table express the probability that the customer number 3 will stock-out before day k when he is scheduled for replenishment on day k.



Day k	$E_i(k)$
1	1.59884
2	1.52965
3	1.46161
4	1.39469
5	1.32886
6	1.26409
7	1.21201
8	1.14921
9	1.07386
10	1.01726
11	.96154
12	.90668
13	.83550
14	.79947
15	.73017
16	.67869
17	.62797
18	.57960
19	.52332
20	.48072
21	.42601
22	.37272
23	.33932
24	.33876
25	.51077
26	1.24010
27	2.95645
28	4.37142
29	7.29664
30	10.42635

Figure 3.3 : The graph and the table present the expected annual incremental cost $E_6(k)$ of replenishing customer number 6 as a function of the scheduled day of replenishment. We also plot on this graph the cost coefficients c_{6k}, for customer number 6, as a function of the day.

CUSTOMER NUMBER 6

Day k	Probability of stocking out before day k
1	.00000
2	.00000
3	.00000
4	.00000
5	.00000
6	.00000
7	.00000
8	.00000
9	.00000
10	.00000
11	.00000
12	.00000
13	.00000
14	.00000
15	.00000
16	.00000
17	.00000
18	.00000
19	.00000
20	.00000
21	.00000
22	.00003
23	.00037
k^*=24	.00262
25	.01283
26	.04514
27	.11949
28	.24817
29	.42047
30	.60356

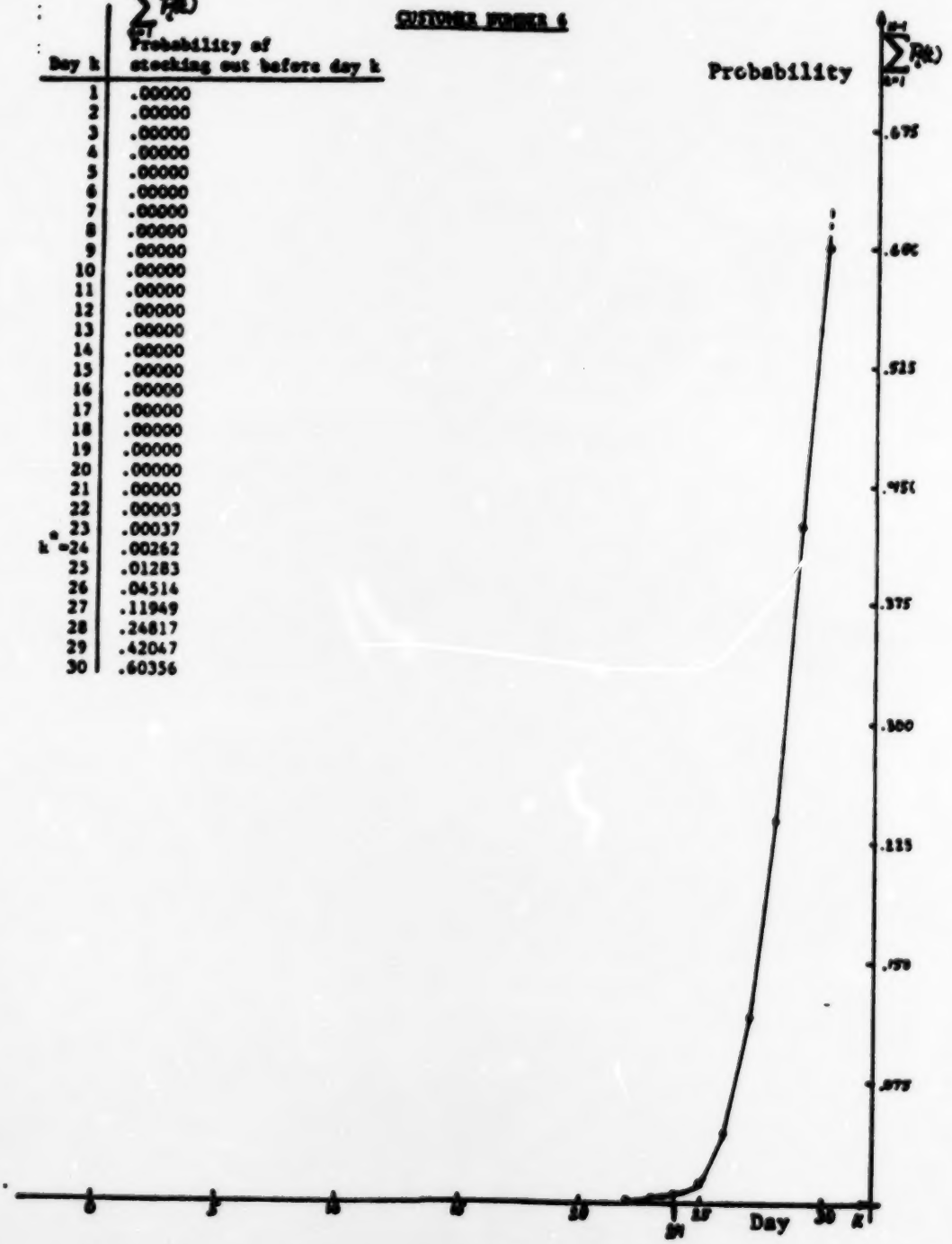


Figure 3.4 : The graph and the table express the probability that customer number 6 will stock-out before day k when he is scheduled for replenishment on day k.

greater than $\sum_{t=1}^{k^*-1} P_i(t)$ and $E_i(k) > E_i(k^*)$. This means that scheduling customer i for replenishment on any day later than day k* increases the expected annual cost. The same is true for any $k < k^*$ with one distinction. The probability that customer i will stock out before day k, i.e. $\sum_{t=1}^{k-1} P_i(k)$, is less than $\sum_{t=1}^{k^*-1} P_i(t)$. The customer selection principle is as follows : Customer i will always be scheduled for replenishment on or before day k*.

In the following paragraph we present the results of the numerical examination of our data base. It must be pointed out that the results rely entirely on the assumption that the customers' daily consumption rates are independent, normally distributed with known parameters and for which their standard deviation is no larger than their mean.

Under those assumptions the numerical examination of the actual data from the York, PA, district, with about 2,200 customers, reveals that : the probability that a customer will run out before the day on which he is expected to reach his safety stock level is less than .05 for 98% of the customers. It is less than .001 for over 90% of the customers. This result is very important for our analysis. If we assign a customer for delivery on a spcific day (on a specific route). the probability that this customer would stock-out some time earlier and consequently need to be replenished by an emergency service is very small, and justifies a deterministic approach to the assignment and routing of the customers. In Figure 4.1 we graph the percentage of customers versus their probability of stocking out before day k* for the data from the York, PA, district.

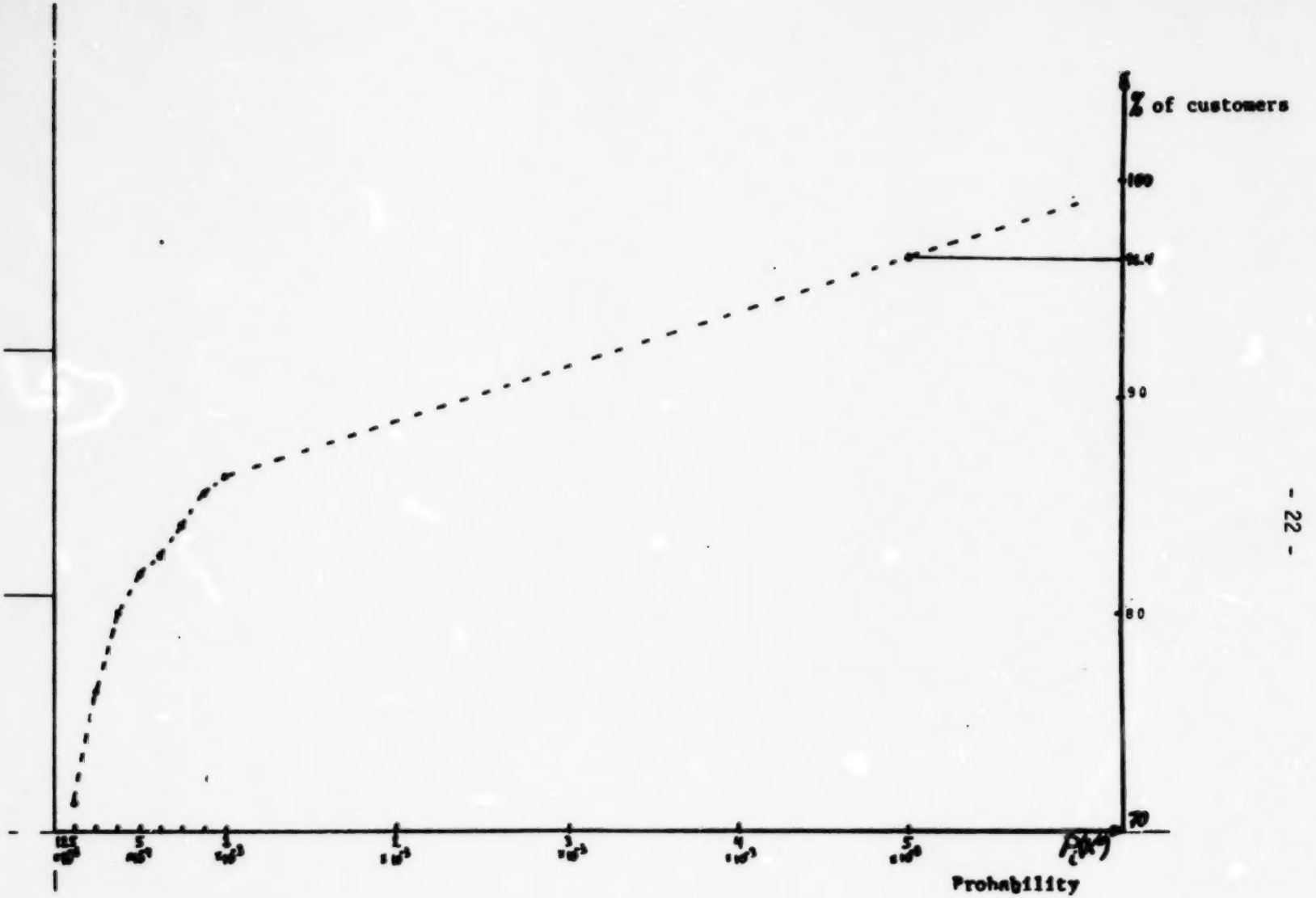


Figure 4.1 : The percentage of customers as a function of the probability of stocking-out before their day k*. (k* for any given customer is the last day on which he can be scheduled for replenishment.)

5. The Actual Customer Selection Procedure

Now we proceed to describe the actual customer selection process : We partition the set of customers N into three disjoint subsets for each planning period. Denote the three subsets by M, M' and N-(M U M'). We assume that we have calculated the safety stock levels for all of the customers in N. The subset M is constructed in the following way. Each customer that is expected to reach his safety stock level during the planning period is included in the subset M. The customers from M must be scheduled for replenishment during the planning period before or on their day k*.

When minimizing the annual distribution cost, we are never sure at what demand level the customer will be replenished in some future planning period even though we know the specific level at which he is being replenished at this time. When we evaluate future distribution costs associated with the replenishment of a customer at a given demand level, our simplifying assumption is that in the future we will consistently replenish this customer when his expected demand matches the specific demand level. This assumption enables us to calculate the expected annual marginal distribution cost for a customer by calculating the expected number of replenishments associated with a demand level. This calculation will most likely result in an overestimation of the cost if we are considering a replenishment to a customer when his demand is low (relative to his tank capacity), and thus tend to substantially penalize the cost of replenishing this customer when his inventory level is high.

For each customer not in the subset M, two factors influence the decision whether or not a customer should be included in the subset M'.

(a) The customer's current inventory level (in other words, the customer's current demand) and

(b) The customer's expected marginal routing cost.

If, for a given customer, his current inventory level is high, then his marginal annual distribution cost is high relative to the cost when his inventory level equals his safety stock level. The difference between the two costs is the incremental marginal annual distribution cost and is represented by the c_{ik} coefficients.

The question in the actual customer selection process for the subset M' is the following : If there would remain some unutilized fleet capacity in the process of assigning customers from the subset M to days in the planning period, which other customers would we consider replenishing? The answer given here is that we would consider replenishing those customers whose incremental marginal annual distribution cost would be less than the cost of maintaining the unutilized part of fleet's capacity. The cost incurred by maintaining a capacity that is not utilized corresponds primarily to the cost of the drivers' idle time. We cannot discuss utilization of a vehicle's capacity if there is no driver to man that vehicle.

In order to construct the subset M' we first need a "good" estimate for the value of the driver's time per delivery, for the case in which the delivery is not required and instead the driver is idle. If, after scheduling all the customers from the subset M for replenishment a driver has a given amount of idle time left, the cost of distribution increases by the cost of this idle time. (That is, the direct distribution cost equals the cost of drivers' productive time, the cost of the fleet's productive time, and the cost of drivers' idle time.) If on a full work day, a driver services on the average D customers, his daily salary is X dollars, and the average vehicle operating cost per delivery is v_a, the a logical candidate to estimate the value of the driver's time per delivery would be $\frac{X}{D} - v_a$. It equals to the average driver's cost per delivery minus the

vehicle's operating cost per delivery which is saved if the vehicle is idle. Denote this cost for the driver's time per delivery a_d. It represents an estimate of the direct cost added to the distribution cost because a driver is idle for a time period sufficient to complete an additional delivery.

Before specifying the procedure that selects the customers in N-M that form the subset M', we present a hypothetical situation in which we describe the logic of the M' selection process.

Assume the following situation : on day k a driver is assigned a set of routes which are not expected to fully utilize his work time. The driver's expected idle time equals the expected marginal service time for customer i ($i \epsilon N-M$).

Just for this case denote the expected marginal cost of servicing customer i by m_i. Assume that on day k the coefficient $c_{ik} = .3m_i$. The interpretation of this value for the c_{ik} is that if in the future customer i will be consistently replenished when his expected demand matches his expected demand on day k, he will require .3 additional deliveries in the course of one year. Assume that if in the future customer i will be replenished consistently when his expected demand matches the expected demand on day k* (i..e., the safety-stock level), he will require a total of ten deliveries annually. (Note that $c_{ik^*} = 0$.) Thus, if customer i will be serviced on day k*, at best, serving him will cost $10\ m_i$.

If the driver's idle time on day k is not utilized, the distribution system incurs the cost estimated by a_d (the cost of driver's time per delivery). If we charge this idle time on day k to customer i, then the annual cost of servicing customer i is $10m_i + a_d$. On the other hand, if we <u>do</u> service customer i on day k, then we estimate the annual service cost as $10.3m_i$. If we ignore the vehicle operating cost, and accept that for

customer i m_i is an average marginal cost of servicing this customer, we can substitute m_i for a_d ; then we compare the $11m_i$ ($10m_i + 1m_i$) incurred by not servicing customer i on day k, with the $10.3m_i$ incurred if we do service him on day k, we must then conclude that we should, if we have the time, service customer i on day k.

Now we are ready to construct the subset M'. If for $i \in N - M$ there exists a day k in the planning period such that $c_{ik} - a_d < 0$ then the i^{th} customer belongs to the subset M' on day k. Note that if $c_{ik'} - a_d < 0$ for some k', then $c_{ik} - a_d < 0$ for all $k > k'$.

We define the cost coefficients for the customers in M' as follows :

$$g_{ik} \equiv c_{ik} - a_d$$

(For all $i \in M'$, the coefficients g_{ik} are negative.)

In the procedure outlined so far for constructing the subset M', we implicitly assume that the locations of the customers are uniformly distributed over some bounded area. The uniform distribution of customers' locations justifies the fact that we use a constant a_d for all the customers (i.e., a_d is customer independent). In the case where customers' locations are not uniformly distributed, we may require an estimate of driver's average cost per customer as a function of the customer's location. In this case, the term a_d will be a function of i and the notation will change to $a_d(i)$. The coefficients g_{ik} will then be calculated as $c_{ik} - a_d(i)$.

We present computational results for both cases ; i.e., in Case 1 we asume that the locations of the customers are uniformly distributed over some bounded area and we compute the constant a_d as the average of the marginal costs of arriving at customers' locations. The resulting constant for our data base has the value of $4.70. The subset M' in Case 1 consists

of all customers i such that c_{ik} - 4.70 < 0 for some k. The g_{ik}'s equal to c_{ik} - 4.70 whenever this expression is negative. In Case 2 the subset M' consists of all customers i such that $c_{ik} - a_d(i) < 0$ for some k and the g_{ik}'s are computed similarly.

Starting from the same initial position, implementing each case resulted in a different subset M' for the first planning period and the next eleven successive planning periods for which we tested our procedures. After selecting the customers for the subsets M and M', in both cases exactly the same routing procedures were implemented to obtain a routing solution for each week.

The measure used in order to evaluate the efficiency of the distribution solution provided for each planning period is expressed in the number of inventory units delivered per hour of distribution operation. (This includes the routing times, the unloading times and the fixed times per delivery.) This measure for the efficiency of the distribution operation is widely accepted in the heating oil delivery industry. It is considered a more objective measure than a cost based measure because cost figures usualy depend on many local and accounting factors. In addition, a solution to the inventory routing problem is not examined only for a single planning period, but for twelve dependent successively-solved weekly problems. Different solutions to this problem start with the same initial data input ; however, for each period it is most probable that different specific inventory problems will be solved. Consequently, a direct cost comparison would no longer be appropriate. Nonetheless, the results obtained using this measure should be viewed with some degree of caution since it does not express the amount (or the cost) of drivers' slack time in the system. In particular, it implicitly assumes that if driver time is reduced then costs can be reduced either by reducing total driver pay hours or by employing

drivers in other areas of the business. It should be up to the district manager to evaluate alternative solutions since different solutions might affect operations other than distribution.

The results are presented in Table I. For Case 1 of this customer selection procedure we also performed some sensitivity tests on the falue of a_d. Under Run I we present the results obtained for $a_d = 3.7$. Under Run II we present the original results (i.e., $a_d = 4.7$), and under Run III, the results for $a_d = 5.7$. Run IV corresponds to Case 2, where the parameter a_d is customer dependent (i.e., $a_d(i)$). The computational results in Table I indicate that the customer selection procedure labled Case 2 has the "best" solution for the inventory routing problem when tested for twelve suiccessive weeks on our data base. The efficiency of this solution is 80 units/hour better than the one presented under Run II. The sensitivity analysis on the parameter a_d for the procedure labled Case 1 indicates that computing a_d as the average of this estimated marginal costs of arriving at a customers' locations (i.e., $a_d = 4.7$) produces a "good" quality solution to the problem. Run II ($a_d = 4.7$) produces a better solution than the other two runs (Run I ($a_d = 3.7$) and Run III ($a_d = 5.7$).

In order to enable the reader to relate to the numbers presented in Table I we note that the distribution company from which the actual data was obtained currently operates at a level of efficiency of about 4000 units/hour. (For more detailed evaluation of the results obtained in this study see [4, 5].)

Partitioning customers into the three disjoint subsets M, M', and N - M U M', divides them into one subset that will be replenished during the planning period-M, a subset that might be replenished during the planning period-M' and a subset that will not be replenished during the planning period-N - MUM. Operationally, this partition is very useful to a

Table I : Units/hour

Week	Run I (a_d=3.7)	Run II (a_d=4.7)	Run III (a_d=5.7)	Run IV (a_d(1)
1	5997	6463	6475	5873
2	6767	7215	7294	7081
3	6443	6645	6599	7120
4	6464	5896	5846	6084
5	6577	6687	6666	6752
6	7129	6996	7137	7196
7	6851	7251	7116	7303
8	6878	6864	6862	7173
9	6263	6432	6030	6561
10	5810	6088	6135	6078
11	6485	6372	5762	6281
12	6373	6512	6849	6641
Units/ hour for the 12 weeks run	6504	6606	6552	6686

dispatcher, even if he does not design the vehicle routes by a computerized routing system. For each planning period the dispatcher can be given a list of customers that must be replenished (M) and another list of customers that can be incorporated into the routes if time permits (M'). If, after a set of specific routes has been designed, a vehicles breaks down, the first customers that the dispatcher can eliminate from the routes are customers from M'. Thus, the partition allows for the addition and deletion of customers from the routes according to the actual realization of the delivery plan.

6. The Alternative Customer Selection Procedure

In the actual customer selection procedure we construct the subset M' of customers by examining the relation between the cost coefficients c_{ik} for each i and the k's which correspond to the days of the given

planning period, as compared with the cost of driver's idle time - a_d (or $a_d(i)$). In the alternative customer selection procedure we construct the subset M' of customers without refering to the cost of the driver's time.

For a given planning period, and a subset M of customers, the construction of M' follows the steps below :

Step 1. Set initially $M' = \emptyset$.

Step 2. Denote by $\overline{L_i} = \min_{p,s \in MUM'} \{V_{pi}+V_{is}-V_{ps}\}$

for $i \in MUM'$

V_{ps} denotes the cost of going from customer p to customer s.

Step 3. Denote by f_{ik} for each $i \in M$ the fraction time remaining since the last replenishment until customer i is expected to reach his safety stock level.

Test for all $i \in MUM'$ if $\overline{L_i} + f_{ik}L_i - L_i$ 0 for some k in the planning period. (L_i represents the estimated incremental routing cost for customer i.)

In the case when this inequality does not hold we stop with the subset M' constructer so far.

In the case when the inequality does hold for some $i \in MUM'$ then to M' add that i for which the inequality has the most negative value and recompute the $\overline{L_i}$ terms for all $i \in M'$. Then return to Step 2.

The coefficients g_{ik}, which express a saving associated with replenishing a customer from M' in the current planning period, are computed as follows :

$$g_{ik} = \text{Min } \{\overline{L_i}+f_{ik}L_i-L_i \quad k \text{ is a day in the planning period}\}$$

The logic behind this alternative customer selection procedure is that it might be very cost efficient (in terms of a particular geography of

the customers already being replenished) to include a particular customer in the routing solution for the given planning period.

The lower bound on the cost of inserting a customer into the routing solution is expressed by $\overline{L_i}$. The estimated incremental routing cost for customer i is represented by L_i. We add i to M' in the case when the estimated additional cost of early replenishment plus the lower bound on the routing cost of the present replenishment are still below the estimated incremental routing cost for i.

This alternative customer selection procedure does not depend on the cost of a driver's idle time. However it is more costly computationally and most important it does not insure the construction of M' even when it might be cost beneficial to do so. The reason is that we construct the subset M' by inserting the customers that do not belong to M one by one according to a specific criterion. It is quite plausible that in some cases when we test each individual $i \varepsilon M$ to see if it should be in M', the answer will be negative ; however if some subset of the customers is already in M' then the answer for all of them is a positive one. We did not test the alternative customer selection procedure computationally on our data, thus its applicability remains an open question.

APPENDIX I

Procedure for Estimating L_i -- The Cost of Arriving at Customer i's location

A naive approach for estimating L_i is to take the cost of one way travel between the depot and the i^{th} customer. This estimate assigns the same cost to customers of equal distance from the depot regardless of proximity of the customer to other customers. In high density geographic clusters of customers this will most likely result in an overestimate of the real cost and in large sparse areas it will underestimate the cost of reaching a customer. We describe a number of intuitive procedures that incorporate some measure of the customers' geographic concentration in their estimate.

Procedure 1. Calculate the average distance of say the S closest customers for each customer in the entire set. Find the ratio of this average distance to the distance between the depot and the customer. Use the ratio as a multiplication factor for the cost of one way travel between the customer and the depot in order to arrive at the estimate for L_i. The parameter S can be modified easily.

Procedure 2. Find S customers closest to customer i. Index them $i_1, i_2, \ldots, i_S$, then calculate the estimate to L_i by the following formula :

$$L_i = \frac{v_{i_1,i} + v_{i_2,i} + \ldots + v_{i_S,i} + v_{01}}{S + 1} .$$

Where $v_{i_j i}$ is the cost of traveling directly between the customer i_j and customer i and v_{0i} is the cost of traveling from the depot to customer i. The logic behind this estimate is that there exists a number S such that in an optimal solution to the inventory routing problem that visits the i^{th} customer, the probability of going directly between the i^{th} and i_j^{th} customer in a cluster of size S is approximately the same as the probability of going directly from (or to) the depot and the i^{th} customer. Thus the v_{0i} should be included in the estimate of L_i. Procedure 2 was the one used when the appropriate L_i was calculated to compute the cost coefficients c_{ik} in Figure 2.1 and Figure 2.3. In that example S = 25 was selected.

Procedure 3. Choose a number of past planning periods, ay p. For each planning period record the routes in a daily list. For each day calculate the cost for the cheapest insertion of customer i into a route. Denote by L_i the average cost for cheapest insertion of customer i into a route.

A slightly different version of this procedure is developed when we record the cost of the cheapest and second cheapest insertion of customer i into a daily routing solution. We calculate the average cost for a cheapest insertion. Denote this cost by L_i^1. Calculate the average cost of the second cheapest insertion. Denote by this cost by L_i^2. L_i can be estimated using a weignted combination of L_i^1 and L_i^2, i.e., $L_i = \alpha L_i^1 + (1-\alpha)L_i^2$ for some $\alpha \in [0,1]$.

REFERENCES

[1] ASSAD, A., GOLDEN, B., DAHL, R. and DROR, M. "Design of an Inventory/Routing System for a Large Propane-Distribution Firm". Proceedings of Southeastern TIMS Meeting, 315-320, October 1982.

[2] ASSAD, A., GOLDEN, B., DAHL, R. and DROR, M. "Evaluating the Effectiveness of an Integrated System for Fuel Delivery". Proceedings of Southeastern TIMS Meeting (1983).

[3] BROWN, R.G., Smoothing, Forecasting and Prediction of Discrete Time Series, Prentice Hall, Inc. (1963).

[4] DROR, M. "The Inventory Routing Problem", Ph.D. Thesis, College of Business and Management, University of Maryland (1983).

[5] DROR, M., BALL, M., GOLDEN, B., BODIN, L. "Inventory Routing Problem". Submitted for publication.

[6] FEDERGRUEN, A. and ZIPKIN, P. "A Combined Vehicle Routing and Inventory Allocation Problem", Research Working Paper No. 345A, Graduate School of Business, Columbia University (June 1980).

[7] FISHER, M., JAIKUMAR, R. and BELL, W. "The Impact of Advanced Technologies on Inventory Management", Proceedings of th Annual Meeting NCPDM (October 1981).

[8] FRANCIS, V.J. "On the Distribution of the Sum of n Sample Values Drawn from a Truncated Normal Population", Journal of the Royal Statistical Society, Series B, 8, 223-232 (1946).

[9] GOLDEN, B., ASSAD, A., DROR, M. and DAHL, R. "Planning a Distribution System in th Presence of Routing and Inventory Subproblems", Paper presented in 1982 Joint National Meeting ORSA/TIMS, San Diego.

[10] LAAKSO, C.A., "Safety Stocks and Service Levels for Multi-Location Inventory. Unpublished M.S. thesis, Illinois Institute of Technology, Chicago, IL. (1978).

[11] THOMPOULOS, N.T. Applied Forecasting Methods, Prentice-Hall (1980).